Chapter 17 - Genomes

OVERVIEW

Chapter 17 explains DNA sequencing, and introduces such concepts as BACs, genomic libraries, and shotgun sequencing. The kinds of information gleaned from sequencing are described, including specific detail on a number of pathogens that may lead to new vaccines and antibiotics. Genomes of different model organisms are compared, including such features as repetitive DNA sequences and transposons. The human genome is described along with such topics as haplotyping and personalized medicine. Chapter 17 concludes with further advances in systems biology known as proteomics and metabolomics.

KEY CONCEPTS

1. Genome sequencing is accomplished by first determining the sequences of small DNA fragments and then aligning them.

2. Bacterial artificial chromosomes (BACs) and genomic libraries are part of the hierarchical strategy of genome sequencing.

3. Shotgun sequencing provides an alternative and faster approach to sequencing that relies on computer-based alignment of many smaller DNA fragments.

4. The use of automated reagent-delivery robots and automated detection of fluorescent nucleotides has greatly increased the speed and efficiency of genome sequencing.

5. Transposable elements and transposons move DNA sequences around within a chromosome.

6. Sequencing the genomes of pathogens has led to insights into their molecular composition; such information may lead to new vaccines and treatments.

7. Shotgun sequencing and metagenomics has led to the discovery of countless new organisms.

8. Much of our knowledge of the eukaryotic genome comes from the study of model organisms such as yeast, the nematode *Caenorhabditis elegans*, the fruitfly *Drosophila melanogaster*, and the mustard *Arabidopsis thaliana*.

9. In comparison to prokaryotes, single-cell eukaryotes (such as yeast) have many new genes targeting proteins to organelles, while multicellular organisms (such as *C. elegans*) have additional genes specifying the structure and function of specialized tissues.

10. Gene duplication has led to many “related” genes within larger genomes. This includes families of genes with related functions as well as apparently nonfunctional pseudogenes.

11. Highly repetitive DNA sequences are present in thousands of copies, either in tandem or scattered in clusters around the genome. They are not transcribed and their function is unknown.

12. Other types of repetitive sequences include certain rRNA coding genes and a variety of transposable elements. Many of these repetitive sequences are thought to be nucleotide parasites lacking in any beneficial function for the cell.

13. The human genome is made up of about 24,000 genes that vary greatly in size. Many contain multiple introns and a large number apparently code for multiple proteins. They are not uniformly distributed throughout the 23 human chromosomes.

14. Single nucleotide polymorphisms (SNPs) define the haplotype of individuals and provide much information about their genetic makeup.

15. Cheaper genotyping may usher in an era of personalized medicine, but epigenetic and environmental effects, as well as complex gene interactions, pose challenges to fulfilling this promise.

16. Mass spectrometry and two-dimensional gel electrophoresis are being employed in efforts to characterize the set of all expressed proteins (i.e., the proteome).

17. The proteome is more complex than the genome because one gene may encode for several different proteins. Differences between organisms ultimately depend on differences in their proteome over their lifespan.

18. The metabolome is the description of all small metabolites in the cell including common primary metabolites and additional secondary metabolites that may be species specific.

KEY TERMS

bacterial artificial chromosome (BAC) A DNA cloning vector used in bacteria that can carry up to 150,000 base pairs of foreign DNA.
bioinformatics The use of computers and/or mathematics to analyze complex biological information, such as DNA sequences.
comparative genomics Computer-aided comparison of DNA sequences between different organisms to reveal genes with related functions.
Functional genomics The assignment of functional roles to the proteins encoded by genes identified by sequencing entire genomes.
gene family A set of similar genes derived from a single parent gene; need not be on the same chromosomes. The vertebrate globin genes constitute a classic example of a gene family.
genomic library All of the cloned DNA fragments generated by the action of a restriction endonuclease on a genome.
haplotype Linked nucleotide sequences that are usually inherited as a unit (as a "sentence" rather than as individual "words").

hierarchical sequencing An approach to DNA sequencing in which genetic markers are mapped and DNA sequences are aligned by matching overlapping sites of known sequence. (Contrast with **shotgun sequencing**.)
high-throughput sequencing Rapid DNA sequencing on a micro scale in which many fragments of DNA are sequenced in parallel.
highly repetitive sequences Short (less than 100 bp), nontranscribed DNA sequences, repeated thousands of times in tandem arrangements.
Human Genome Project An effort to determine the DNA sequence of the entire human genome, understand the structures and functions of the genes, make comparisons with other organisms, and understand the social implications of this information.
metabolome The quantitative description of all the small molecules in a cell or organism.
metagenomics The practice of analyzing DNA from environmental samples without isolating intact organisms.
moderately repetitive sequences DNA sequences repeated 10–1,000 times in the eukaryotic genome. They include the genes that code for rRNAs and tRNAs, as well as the DNA in telomeres.
pharmacogenomics The relaionship between an individual's genetic makeup and response to drugs.
proteome The set of proteins that can be made by an organism. Because of alternative splicing of pre-mRNA, the number of proteins that can be made is usually much larger than the number of protein-coding genes present in the organism's genome.
pseudogene [Gk. *pseudes*: false] A DNA segment that is homologous to a functional gene but is not expressed because of changes to its sequence or changes to its location in the genome.
shotgun sequencing A relatively rapid method of DNA sequencing in which a DNA molecule is broken up into overlapping fragments, each fragment is sequenced, and high-speed computers analyze and realign the fragments. (Contrast with **hierarchical sequencing**.)
transposable element A segment of DNA that can move to, or give rise to copies at, another locus on the same or a different chromosome.
transposon Mobile DNA segment that can insert into a chromosome and cause genetic change.